

# Basic Probability Notes

Ramesh Sridharan\*

These notes give a review of basic probability, focusing on a few important concepts. For a more thorough treatment, see any introductory probability book; I recommend *Introduction to Probability* by Bertsekas & Tsitsiklis.

## Probability review

Here's our setup: we have *experiments* which generate *outcomes*. The *sample space* is the set of all outcomes of an experiment. *Events* are subsets of the sample space (which can possibly be empty). A *probability distribution* assigns probabilities to events, and must satisfy the three axioms of nonnegativity, additivity, and normalization.

*Conditioning* on an event means we have observed that event occurring. This gives a new probability distribution; it can be thought of as restricting the universe of possibilities to those outcomes which satisfy the condition.

Two events  $A$  and  $B$  are defined to be *independent* if  $\mathbb{P}(A|B) = \mathbb{P}(A)$ , or equivalently if  $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$ . Intuitively, two events are independent if knowing one does not change the probability of the other one occurring.

*Bayes' rule* states that

$$\mathbb{P}(A|B) = \frac{\overbrace{\mathbb{P}(B|A)}^{\text{"likelihood"}} \overbrace{\mathbb{P}(A)}^{\text{"prior"}}}{\mathbb{P}(B)}.$$

Typically,  $B$  is an observed event and  $A$  is an event representing the hidden state of the world (so  $P(A|B)$  is our belief about the hidden state of the world given our observations). Our problem-specific model will provide likelihood and prior distributions  $P(B|A)$  and  $P(A)$  respectively. Computing the denominator often requires the use of total probability (or, in the case of random variables, marginalization), and can sometimes be computationally intractable.

The law of *total probability* states that given events  $A_1, \dots, A_n$  that are disjoint (i.e., they don't overlap) and exhaustive (i.e., they cover the entire sample space when combined), we can write

$$\mathbb{P}(B) = \sum_{i=1}^n \mathbb{P}(B|A_i)\mathbb{P}(A_i)$$

This is useful when  $\mathbb{P}(B)$  is difficult to compute, but  $\mathbb{P}(B|A_i)$  is easier.

Also recall that the notation  $A^c$  refers to the complement of an event.

---

\*Contact: rameshvs@csail.mit.edu

## Basic combinatorics

Suppose we wish to form a 4-letter string using the letters of the alphabet. Suppose we require that the letters must be different (i.e. we sample them *without replacement*). Then, there are 26 possibilities for the first letter, and for each of those, there are 25 possibilities for the second letter, and so on. So, the number of possible strings is given by  $26 \cdot 25 \cdot 24 \cdot 23$ .

In general, a length- $k$  **permutation** of  $n$  elements is an ordered collection of  $k$  of the  $n$  elements. There are  $\frac{n!}{(n-k)!}$  different ways to construct such a permutation. In this example,  $n = 26$  letters of the alphabet and  $k = 4$  letters, so the formula gives us  $26!/22! = 26 \cdot 25 \cdot 24 \cdot 23$ .

When we are not concerned with order, we can simply take the number of permutations and divide by the number of ways to reorder the  $k$  elements,  $k!$ . This gives  $\frac{n!}{k!(n-k)!}$ .

In our earlier example, if we didn't care about order, we would divide by the number of ways to rearrange the 4 letters (which is  $4! = 4 \cdot 3 \cdot 2$ ). Our final answer would then be  $\frac{26!}{22! \cdot 4!} = \frac{26 \cdot 25 \cdot 24 \cdot 23}{4 \cdot 3 \cdot 2}$ .

# Problems

## Example 1

A family has two children. For each piece of information below, determine the probability that both children are boys given that piece of information. For this problem, assume that every child is equally likely to be a boy or a girl, and the genders of the two children are independent.

- (a) No information (i.e. compute the unconditional probability that both children are boys).
- (b) The younger child is a boy.
- (c) At least one child is a boy.
- (d) At least one child is a boy born on a Tuesday. Assume each child is equally likely to be born on any day of the week.

**Solution:** This problem was posed by Gary Foshee at a gathering of puzzle enthusiasts. You can find lots of discussion about it online <sup>1</sup>.

- (a) This is simply the probability of two independent events, each of which occurs with probability  $1/2$ . Therefore, the probability is  $\frac{1}{4}$ .
- (b) Carefully examining the sample space, we see that there are 4 possible outcomes:

First child	Second child
<b>Boy</b>	<b>Boy</b>
<b>Boy</b>	<b>Girl</b>
Girl	Boy
Girl	Girl

Using computations similar to those in (a), we can see that all 4 possibilities are equally likely. If we observe that the younger child is a boy, we can eliminate the third and fourth possibilities. One out of the two remaining possibilities satisfies the event we are looking for (“both children are boys”). So, the probability is  $\frac{1}{2}$ .

- (c) Once again, we can write out the entire sample space:

First child	Second child
<b>Boy</b>	<b>Boy</b>
<b>Boy</b>	<b>Girl</b>
<b>Girl</b>	<b>Boy</b>
Girl	Girl

Here, our conditioning event, “at least one child is a boy,” only eliminates one possibility, leaving 3 equally likely candidates. Once again, only the first out of the three satisfies the event we’re looking for. So, the probability is  $\frac{1}{3}$ .

---

<sup>1</sup>See the New Scientist article or a useful blog post

- (d) In order to write out the full sample space for this problem, we'd need 196 entries: there are 14 possibilities for the first child, and for each of those, 14 possibilities for the second. So, we'll solve a simpler problem: suppose babies can only be born on Tuesdays and Wednesdays. Using B, G, T, and W to abbreviate boy, girl, Tuesday and Wednesday respectively, the sample space is:

First child	Second child
<b>B,T</b>	<b>B,T</b>
<b>B,T</b>	<b>B,W</b>
<b>B,T</b>	<b>G,T</b>
<b>B,T</b>	<b>G,W</b>
<b>B,W</b>	<b>B,T</b>
B,W	B,W
B,W	G,T
B,W	G,W
<b>G,T</b>	<b>B,T</b>
G,T	B,W
G,T	G,T
G,T	G,W
<b>G,W</b>	<b>B,T</b>
G,W	B,W
G,W	G,T
G,W	G,W

The 7 bolded entries satisfy the event we're conditioning on: that one of the children is a boy born on a Tuesday. Out of these, only 3 (the first two and the fifth) correspond to two-boy families. Therefore, the probability is  $\frac{3}{7}$ .

While the sample space for the full problem is too large to write out easily, we can extend this reasoning to obtain the final answer  $\frac{13}{27}$ . I encourage you to try arriving at this answer for yourself!

In case you don't believe the answer (as I didn't the first time I heard this puzzle), one thing you can always do is to simulate it. The following python code (also available from the same place you found these notes) simulates this problem:

---

```

1 from __future__ import division
2
3 import sys
4 import random
5
6 def generate_families(N_families, N_children_per_family):
7     """
8     Generates a list of two-child families, where each family is a two-element
9     list of (birth day of week, gender) pairs.
```

```

10     """
11     all_families = []
12     days = ['Sunday', 'Monday', 'Tuesday', 'Wednesday',
13            'Thursday', 'Friday', 'Saturday']
14     genders = ['boy', 'girl']
15     for ii in xrange(N_families):
16         family = []
17         for child in xrange(N_children_per_family):
18             day = random.choice(days)
19             gender = random.choice(genders)
20             family.append((day,gender))
21         all_families.append(family)
22     return all_families
23
24 def test_condition(family):
25     """ Tests whether a family has a boy born on Tuesday. """
26     for child in family:
27         (day, gender) = child
28         if day == 'Tuesday' and gender == 'boy':
29             return True
30     return False
31
32 def test_two_boys(family):
33     """ Tests whether a family has two boys. """
34     ((day1, gender1), (day2, gender2)) = family
35     return (gender1 == gender2 == 'boy')
36
37 def example_problem(families):
38     """
39     Given a list of families, returns the conditional probability
40     of having two boys given that one is a boy born on a Tuesday.
41     """
42     conditioned = [fam for fam in families if test_condition(fam)]
43     fully_satisfied = [fam for fam in conditioned if test_two_boys(fam)]
44
45     conditional_probability = len(fully_satisfied) / len(conditioned)
46
47     print("Conditional probability was %f."%conditional_probability)
48
49 if __name__ == '__main__':
50     N_families = int(sys.argv[1])
51     families = generate_families(N_families, 2)
52     example_problem(families)

```

---

One way to view these problems is on an “information spectrum.” In each situation, we are

given some information about the family, and asked for the probability of having two boys. This probability was smallest in (a), where we were given no information. As we are given more and more information, the probability of this event increases: part (c) provides a small amount of information, part (d) provides a little more, and part (b) provides even more still. Our smaller problem in part (d) doesn't provide as much information as the full problem: we only eliminate half the possible days the child could have been born on. However, the full problem eliminates 6/7 of the days by telling us that the boy was born on a Tuesday.

### Example 2

You have an urgent assignment due in only a few minutes. You know it's in one of your drawers, but you're not sure which one. The probability that the assignment is in drawer  $k$  is  $D_k$ . If drawer  $k$  has the assignment and you search there, you have probability  $p_k$  of finding it.

Suppose you search drawer  $i$  and do not find the assignment. Like any good student, you decide to spend your little remaining time computing probabilities:

- (a) Find the probability that the paper is in drawer  $j$ , where  $j \neq i$ .
- (b) Find the probability that the paper is in drawer  $i$ .

**Solution:** Let  $A_k$  be the event that the assignment is in drawer  $k$ , and  $B_k$  be the event that you find the assignment in drawer  $k$ .

- (a) We'll express the desired probability as  $\mathbb{P}(A_j|B_i^c)$ . Since this quantity is difficult to reason about directly, we'll use Bayes' rule:

$$\mathbb{P}(A_j|B_i^c) = \frac{\mathbb{P}(B_i^c|A_j)\mathbb{P}(A_j)}{\mathbb{P}(B_i^c)}$$

The first probability in the fraction,  $\mathbb{P}(B_i^c|A_j)$ , expresses the probability of not finding the assignment in drawer  $i$  given that it's in a different drawer  $j$ . Since it's impossible to find the paper in a drawer it isn't in, this is just 1.

The second probability,  $\mathbb{P}(A_j)$ , is given to us in the problem statement as  $D_j$ .

The third probability,  $\mathbb{P}(B_i^c) = 1 - \mathbb{P}(B_i)$ , is difficult to reason about directly. But, if we knew whether or not the paper was in the drawer, it would become easier. So, we'll use total probability:

$$\begin{aligned} \mathbb{P}(B_i) &= \overbrace{\mathbb{P}(B_i|A_i)}^{\text{prob. of finding in } i \text{ if in } i} \underbrace{\mathbb{P}(A_i)}_{\text{prob. of being in } i} + \overbrace{\mathbb{P}(B_i|A_i^c)}^{\text{prob. of finding in } i \text{ if not in } i} \underbrace{\mathbb{P}(A_i^c)}_{\text{prob. of not being in } i} \\ &= p_i D_i + 0(1 - D_i) \end{aligned}$$

Putting these terms together, we find that

$$\mathbb{P}(A_j|B_i^c) = \frac{D_j}{1 - p_i D_i}$$

(b) Similarly, we'll use Bayes' rule:

$$\mathbb{P}(A_i|B_i^c) = \frac{\mathbb{P}(B_i^c|A_i)\mathbb{P}(A_i)}{\mathbb{P}(B_i^c)} = \frac{(1-p_i)D_i}{1-p_iD_i}$$

### Example 3

Suppose you have  $n$  ball-filled urns, numbered 1 through  $n$ . Urn  $i$  has exactly  $i - 1$  green balls and  $n - i$  red balls.

- (a) You choose an urn uniformly at random, and draw a ball without looking at it. What is the probability that the ball you drew is green?
- (b) You then look at the ball and see that it is green. You draw another ball from the same urn. What is the probability that the second ball is green?

**Solution:**

- (a) Let  $N_{\text{red}}$  be the total number of red balls, and  $N_{\text{green}}$  be the total number of green balls.

$$N_{\text{red}} = \sum_{i=1}^n (i-1) = \frac{1}{2}n(n+1) - n = n^2 - n,$$
$$N_{\text{green}} = \sum_{i=1}^n (n-i) = n^2 - \frac{1}{2}n(n+1) = n^2 - n,$$

where the summations come from the formula  $\sum_{i=1}^n i = \frac{1}{2}n(n+1)$ .

Since the urns all have an equal number of balls, randomly choosing an urn and then randomly choosing a ball is equivalent to randomly choosing a ball from all the balls (convince yourself that this is true!). So, the probability is  $1/2$ . This answer should also make intuitive sense: since red and green are completely symmetric here, the answer should be the same for red as it is for green: the only way for that to be true is for both to be  $1/2$ .

- (b) This question would be much easier if we knew which urn the balls came from. So, we'll use total probability to condition on that information. Let  $G_k$  be the event that the  $k$ th ball drawn is green, and  $U_i$  be the event that they were drawn from the  $i$ th urn. Then

the  $U_i$ s are disjoint, exhaustive events:

$$\begin{aligned} \mathbb{P}(G_2|G_1) &= \frac{\mathbb{P}(G_2 \cap G_1)}{\mathbb{P}(G_1)} \\ \text{(total probability)} &= \frac{1}{(1/2)} \sum_{i=1}^n \mathbb{P}(G_2 \cap G_1|U_i) \overbrace{\mathbb{P}(U_i)}^{1/n} \\ \text{(Chain rule)} &= \frac{1}{(1/2)} \sum_{i=1}^n \mathbb{P}(G_1|U_i) \mathbb{P}(G_2|G_1 \cap U_i) \cdot \frac{1}{n} \\ &= \frac{1}{(1/2)} \sum_{i=1}^n \frac{i-1}{n-1} \cdot \frac{i-2}{n-2} \cdot \frac{1}{n} \\ &= \frac{1}{(1/2)} \cdot \frac{1}{3} = \frac{2}{3}, \end{aligned}$$

where the summations can be computed using a computer algebra system such as Wolfram Alpha.

The key to the solution was to see that we didn't have any way of directing computing probabilities of events involving  $G_2$  and  $G_1$ , but once we conditioned on  $U_i$  (i.e., once we knew which urn the balls came from), we could fully use the information we were given.

Note that our answer is quite a bit higher than  $1/2$  even though we've already taken away a green ball from our chosen urn. Intuitively, this is because drawing a green ball first tells us that we are more likely to have chosen an urn with a higher concentration of green balls.